A New Classification of Information: A Step on the Road to Interpretability

Larry H. Reeker
Information Technology Laboratory
National Institute of Standards and Technology
Gaithersburg, Maryland
larry.reeker@nist.gov

Albert T. Jones
Manufacturing Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, Maryland
albert.jones@nist.gov

Abstract

Complex systems, such as manufacturing supply chains, are often modeled as a collection of interacting components with information flows between them. These components are frequently responsible for making a wide range of decisions that are implemented using an optimization, heuristic, or control technique. The traditional approach to system performance focuses on the performance of these components. The view has been that to improve the system performance one had only to develop better techniques. In this paper, we argue that inadequate attention has been paid to the relationship between information and system performance.

Information has played an important role in the manufacturing systems of the past. It will play a dominant role in the Internet-based manufacturing systems of the future. To better design, engineer, implement, and control these systems, we need a fundamental understanding of information and its effects on system dynamics. This paper contends that we need a new characterization of information, a delineation of its salient properties, quantitative metrics for those properties, methods for computing these metrics, and linkages between these metrics and system performance. We focus principally on the first of these, a new characterization of information, and discuss the implications of suggested characterizations for metrics and their measurement, suggesting some approaches for further research.

Keywords: *chaos; complex system; entropy; information; metrics; system performance; ontologies; satisficing*

1. BACKGROUND

The Internet has made the globalization of manufacturing systems, commonly called supply chains, a reality. This globalization has caused two fundamental transformations in the behavior of these systems. First, the rigid organizational hierarchies, typified by the keiretsu in Japan, have been replaced by more flexible, network-like organizational structures. The tightly integrated, closed relationships of the kieretsu had many advantages for both the original equipment manufacturers (OEMs) and the suppliers. The OEMs had a ready set of local, qualified suppliers who were ready, willing, and able to serve their needs. The suppliers, on the other hand, had a guaranteed customer who provided predictable production and delivery dates. This captive relationship shielded both the OEM and its suppliers from the global marketplace. As history has shown, the impact can be positive for a while; but, over time, this shield will weaken the market position of the OEM and the capabilities of the suppliers. After years of observation and emulation, many manufacturers are attempting to build a business structure that will yield the benefits of the keiretsu, but avoid its weaknesses. In these structures, which are self-organizing and Internet-centric, the suppliers and OEMs form a virtual supply chain. This allows OEMs the freedom to choose the best suppliers and suppliers the opportunity to find other customers.

The second fundamental transformation caused by the Internet involves the roles that OEMs and suppliers play in the supply chain. Both have evolved from systems that are principally producers and consumers of physical objects into systems that are also producers and consumers of informational objects. This evolution has taken place in two distinct phases. During the first phase, the OEMs gradually shifted production of all components and sub-assemblies to independent suppliers. Many of these suppliers were located in other countries, which reduced the direct labor cost but increased the logistics and transportation costs. The suppliers became the builders of components and the OEMs became the final assemblers. During this phase, the business aspects of these relationships, and the information associated with them, were still handled by telephone and paper.

During the second phase, which is still ongoing, the Internet has made it possible to exchange electronically not only design and production information, but also business information. The potential exists to conduct all business transactions over the Web. Demand information, logistics

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **AUG 2002**	2. REPORT TYPE	3. DATES COVERED **00-00-2002 to 00-00-2002**
4. TITLE AND SUBTITLE **A New Classification of Information: A Step on the Road to Interpretability**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National Institute of Standards and Technology (NIST),Manufacturing Engineering Laboratory,100 Bureau Dr,Gaithersburg,MD,20899**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES **Proceedings of the 2002 Performance Metrics for Intelligent Systems Workshop (PerMIS ?02), Gaithersburg, MD on August 13-15, 2002**		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

information, purchase-order information, warehouse information, and so on, can be sent anywhere in the world. The Internet can assure that these informational objects are delivered on time and error free. It cannot assure that supply chain partners will interpret these objects in the same way. Furthermore, as described below, decisions made on the wrong interpretation can have dramatic impacts.

Lee, Padmanabhan, and Wang discussed financial impacts that befell some major companies that made purchasing and production decisions based on a misunderstanding of a variety of information. Hewlett-Packard stockpiled laser printers, worth millions of dollars, in response to --what turned out to be-- phantom orders from resellers. Procter & Gamble saw wild fluctuations in orders from their distributors, although its market research showed that the demand for diapers had remained constant. These are two among several examples described in [Jones *et al* 02]. In each case, the decision maker – software or human – made the decision based on its understanding of all available information and a belief that the markets would, in fact, evolve as predicted. Unfortunately, the understanding was incorrect and the resulting predictions were grossly inaccurate. The authors summarized the generic problem as follows, "Distorted information from one end of the supply chain to the other can lead to excessive inventory investment, poor customer service, lost revenues, ineffective transportation, and missed production schedules."

In supply chains, this phenomenon is called the bullwhip effect because small deviations in customer demand can amplify quickly (whip) through the entire supply chain. As indicated above, these small deviations can lead to dramatic changes in performance. This type of dependence on initial conditions is typical of a special class of non-linear, dynamic systems that are called chaotic. Chaotic systems are a subset of the more general class of complex systems.

In this background section, the references to "information" have assumed that the reader has an intuitive idea of what it is. After all, we talk about "sending information" or "looking for information" regularly, and the public recognizes a whole area of technology as "information technology". The word "information" in all of these areas refers to at least two fundamentally different things. One is a physical aspect of information that allows it to be communicated. That physical dimension is absolutely essential. But there is something that is often more important to the human user, and that is the knowledge that is conveyed by the physical manifestation, and we call that "information" too. We are going to have to differentiate the two types later in the paper; but for now, we will continue to look for any effect that what we commonly call "information" on systems, and on the extent to which the information can be used to predict system behavior. We will start with the physical manifestation, which is at the heart of traditional approaches.

2. TRADITIONAL APPROACHES

Complex systems, such as manufacturing supply chains, are often modeled as a collection of interacting components with information flows between them. These components are frequently responsible for making a wide range of decisions that impact the behavior and performance of the entire system. Early in his landmark book, *The Sciences of the Artificial*, Herbert Simon [69] said that in such systems, "it is the organization of the components, not their physical properties, that determine behavior". We interpret this to mean that information and the ability of components to deal with information have a major impact on system performance. Near the end of his book, Simon argues that complex systems exhibit emergent properties -- "given the properties of the parts, and the laws of their interactions, it is not a trivial matter to infer the properties of the whole".

Simon goes on to say that the evolution of these systems is typically non-linear, often chaotic, and sometimes catastrophic. This means that the actual performance of the system can deviate substantially from the predicted performance. Furthermore, small changes in initial conditions can lead to dramatic changes in the evolution of the system. Before discussing our approach, let us review briefly the traditional approaches to these problems

2.1 Input Characteristics

An input, X, is characterized as either deterministic or non-deterministic, depending on whether its true value is known, or assumed to be known, with certainty or not. A great deal of effort has been spent trying to model non-determinism with probability distributions. In some cases, such as queuing systems, assumptions that lead to specific forms for the distribution -- such as Poisson arrivals and Exponential service -- are often made. In most cases, however, distributions are estimated statistically from sample data. Two approaches have been used: a frequency approach and a Bayesian approach.

2.1.1 Frequency Approach. The frequency approach treats the true value X as an unknown constant. The output of a frequentist statistical analysis is an estimate of the expected value and

standard deviation X. Consider the simple case where X is estimated from a sample of n measurements that are assumed to be independent and identically normally distributed random variables with mean c and variance σ^2. Let x and s^2 denote the sample mean and the sample variance of the n measurements. Then x, s^2, and s are the estimates of μ, σ^2, and σ respectively. The probability distribution of x, called a sampling distribution, is also normal but with expected value μ and variance σ^2/n. The ratio $s/\sqrt{n}$ is an estimate of $\sigma/\sqrt{n}$. The standard deviation $\sigma/\sqrt{n}$, called population standard deviation of the mean, characterizes the tightness of the sampling distribution of x about E(x) = μ. So $s/\sqrt{n}$ is called sample standard deviation of the mean, is a measure of the uncertainty about x as an estimate of μ. Once we have these estimates, we use them, the original data, and a goodness-of-fit test to find the distribution. This approach is very sensitive to the sample size and the underlying normality assumption.

2.1.2 Bayesian Approach. The Bayesian approach starts with a prior probability distribution p(X), which can be found using the principle of maximum entropy [Jaynes 68]. This distribution represents the state of knowledge about X before the data is taken. The expected value, the variance, and the standard deviation of the prior distribution p(X) are denoted by E(X), V(X), and SD(X) respectively. The relationship between the measurement data and X is expressed by a function ϕ(data | X) that is obtained by the rules of probability theory from the probability distributions of individual measurements that depend on X. After measurement data are known, the function ϕ(data | X) may be regarded as a function not of "data" but of X. When so regarded this function is called the "likelihood function" of X for given data and written as L(X | data). Bayes' theorem states that the probability distribution of X after measurement, called the posterior distribution and denoted by p(X | data), is proportional to the product of the L(X | data) and p(X). That is, p(X | data) $\propto$ L(X | data) $\times$ p(X). This new distribution represents the state of knowledge about X after measurement. The expected value, the variance, and the standard deviation of the posterior distribution are denoted by E(X | data), V(X | data), and SD(X | data) respectively. The posterior expected value E(X | data) may be taken as the estimated value of X. And the posterior standard deviation SD(X | data) may be taken as the Bayesian evaluation of uncertainty concerning X after measurement. Unlike the frequency approach, the validity of this approach does not depend on a normality assumption or a large sample size.

2.2 System Evolution

System evolution is described in terms such as continuous/discrete, linear/non-linear, static/dynamic, and deterministic/stochastic. A particular system will be characterized by some combination of these terms. Complex systems such as manufacturing supply chains are composed of many manufacturing enterprises, each of which is a system according to our earlier definition. Each of these enterprises, in turn, is composed of many components, which are also complex systems in their own right. Given this, how can we predict the behavior of the entire supply chain?

There is a growing consensus that many manufacturing systems exhibit chaotic behavior [Herrin 01]. This means that they are (1) non-linear and dynamic, (2) discrete or continuous, and, more importantly, (3) deterministic but subject to stochastic influences. So, although they are deterministic, their performance cannot be predicted with certainty in advance. In fact, small changes to the initial state may cause significant changes in the evolution and the performance of the system.[1] Sometimes these changes are gradual and build up over time (Figure 1a), sometimes they are sudden and lead to instabilities in the system and a dramatic degradation in performance (Figure 1b).

As noted above, the traditional approach to dealing with these problems has been through some type of optimization techniques for individual systems or sub-systems. The characteristics of both the inputs and the system dynamics determine the types of technique that is used. If everything is deterministic and linear, well-known operations research, artificial intelligence, or control theory techniques can be used. If determinism cannot be assumed, then techniques that are more complicated must be used. These techniques include utility theory, stochastic optimization, discrete event simulation, and stochastic control theory. Except in the simplest cases, non-linearity is usually dealt with by using a linear approximation.

Based on Simon's arguments, we believe that focusing only on the techniques used to make decisions will not improve necessarily the performance of the system. We believe that more attention should be focused on the inputs to those decisions, information, which can tell us the nature of the inputs and how they might mesh between components of the system. There are, however, fundamental scientific limitations to our

[1] This is in stark contrast to linear systems, where small changes in the inputs lead only to small changes in the outputs.

understanding of information and its impact on the behavior of complex, systems. In particular, the lack of computer interpretable measures of the meaning of, and associated uncertainties for, information can lead to decisions that create chaotic and unstable behavior in these systems.

3. INFORMATION CHARACTERIZATION

If all the information that is important in characterizing a system's behavior could simply be expressed in bits, it would provide a numerical value that could be used to measure and control system performance and make improvements, which we need for the reasons expressed in the previous sections. In a purely physical system, numerical measures of energy output characteristics alone may

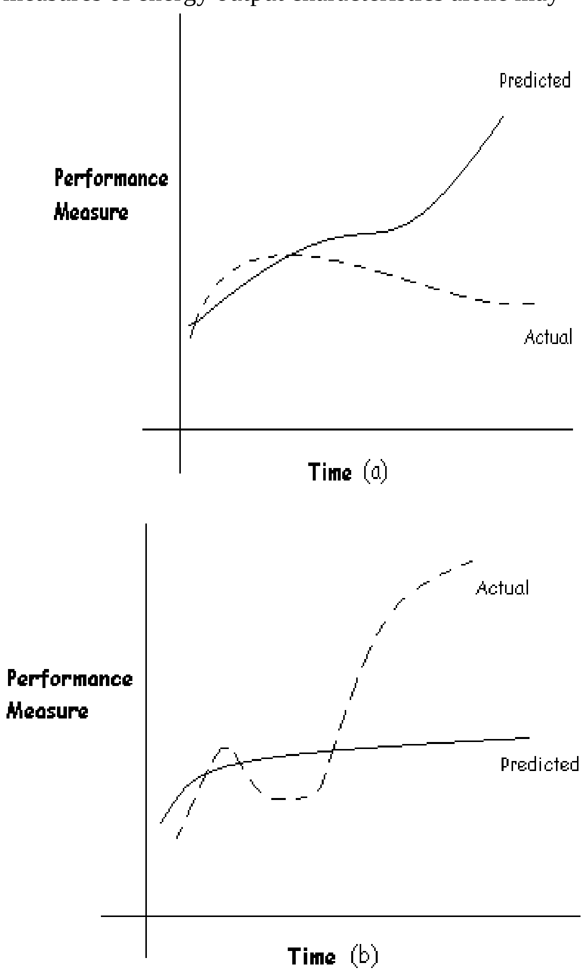

Figure 1. Actual vs Predicted Performance

tell us whether it is sufficient for a particular task and also determine its sufficiency as a part of a larger system. We can also determine energy efficiency. But we still lack adequate measures of information outputs (or internal information for control purposes) from information or hybrid physical- informational systems of any complexity.

The *state* of a system is what we are trying to deal with, whether the system is physical or informational or both. That state is an information object, and the succession of states specifies all the behaviors and the causes of those behaviors. When the state information represents energy or forces, it is expressible simply in numbers or vectors. It tends also to be fairly local in its influence. Even if a system is not strictly a Markov process, it is frequently expressible in such terms. The *physical aspects* of human speech and many other physical phenomena have been fairly well characterized (as indicated by predictive ability) using Markov models.

In informational systems, the models seem inherently to be significantly more difficult. This is true even for natural language syntax, as Chomsky showed: Markov models are not adequate. If one factors in the needed semantics for one person to understand another, it becomes obvious that in human language understanding there is an enormous amount of information that is stored for long periods by a listener and used to understand a speaker at unpredictable times in the future [Miller and Chomsky 63]. One would like to think that industrial supply chains would have more constraints and that the information needed would therefore be more localized; but that is not a foregone conclusion. Standards are one way to apply the needed limits, but standards require better characterization and measurement of parameters. Below we look at some problems with the use of ordinary information theory as we examine the use of system states and their components and communication of knowledge, then look at some other possibilities for characterizing the information in systems.

3.1 Information Within Systems

The definition of a system state is often only an abbreviation of the essential information needed to characterize the system, as indicated in the dictionary definitions:

> **State:** Any of various conditions characterized by definite quantities (as (i.e. of energy, angular momentum, or magnetic moment) in which an atomic system may exist [Merriam-Webster, 2002]

> **State:** The condition of a physical system with regard to phase, form, composition, or structure [American Heritage, Fourth Edition, 2000]

> **State:** The way something is with respect to its main attributes [WordNet ® 1.6, © 1997]

> **State:** How something is ; its configuration, attributes, condition, or information content. The state of a system is usually temporary and volatile. [Free On-line Dictionary of Computing, 2001]

The last of these definitions is closest to what we need to characterizing the information in a system, in that it mentions the information content; of course, the other components it mentions are just more information, but some of them may have physical parameters. In the long run, it is probably best to be inclusive for cases where information plays an essential role and broaden the concept of state (of a system) to the following:

> ***State:*** *All the information at a given instant that is relevant to the behavior of the system at any later time.*

A trace of the system states under all of the conditions in which it will operate is a full informational description of the system. As an example, consider a simple algorithm being executed on a machine – say, a sorting algorithm. It has a series of states that lead from inputs of information (a list to be sorted and an ordering relation) to outputs of the ordered list. There are many well-known sorting algorithms, and each of these, given the same input, will produce a series of states, of which the last state will include knowledge of the initial list and the ordered list. If we compare two such series for a given input and different algorithms, they tell us something about the comparative properties of the algorithm, including efficiency on the particular input (and, by generalization, on whole classes of inputs). These traces may also point out some subtle differences between two algorithms, but if the states contain the same information at a given time, we can assume that they are doing the same thing. That is actually a very strong requirement, of course, since if two had some different information at the same time and did not interact with another system or provide an output at that time, the results would be of only theoretical interest.

In certain cases, it is possible to give a numerical measure to the amount of information (in the Shannon sense) in a state, which can tell us little about what the system is doing. In the sorting case, for instance, if we compare algorithms for the algorithms "quicksort" and "merge sort" by computing amount of uncertainty about the final ordering at each state, we find even though the states have different information at given times, aspects of their behavior are explained by the information measures. For one thing, we can make information-theoretic arguments that show why and when they are most efficient and why they have the same expected time in certain cases. But the particulars of what they are doing and how they differ are not in those figures; so if – for instance – they had to stop after a given time, one might provide a better output than the other.

The simple example, though not very important in these two similar sorting algorithms, illustrates a general problem that we have with the characterization of information as negative entropy (Shannon Information) in predicting system behavior. There are other types of information measures more closely related to computation, such as Chaitin Information and Kolmogoroff Information, but these measures, all based on entropy suffer from similar problems to Shannon Information. In general, the information is incomplete because it does not convey knowledge, but merely a measure of *potential information* in a system. Potential information (amount) is not adequate for understanding of information systems.

3.2 Potential and Mediate Information

Shannon's information has been a highly satisfactory measure of the physical transmission of symbols over a communication channel. Since communicated information always has a physical dimension, the model is relevant, but the physical part is only a carrier for what is actually *meaningful*, and meaningfulness lacks a satisfactory theoretical basis. The physical information that is sent out is not meaningful until it is interpreted when it reaches its recipient. Before that time, it is only data, or "potential information". Shannon himself did not advocate using the term "information", since he pointed out that it did not concern meaning, but the term has stuck.

A diagram may help to illustrate the relationship between potential information and meaningful information (often called "semantic content" or "knowledge"). Modeling it requires a theoretical construct, which will be called herein *mediate information*. That mediate information directs the process by which the potential information becomes meaningful. To provide a graphic example of these constructs and how they interact, Figure 2 includes a version of the model used by Shannon for a generic communication channel. It is labeled for a particular example of potential information (a spoken utterance in a human language going – by sound waves – from one person to another), with some ideas on the relevant mediate information to make that type of potential information meaningful. The successful

transmission of meaningful information in the case of a simple linguistic utterance requires that a certain amount of potential information be transmitted, but it also requires a "hidden channel" of mediate information that is not transmitted with the potential information, but is known previously by the sender and the receiver. It is as if the speaker had encrypted something and sent a message whose key was the mediate information sent by another channel.

The mediate information needed to convert the potential information to meaningful information in Figure 2[2] (and generally) is partly a set of conventions that had been used by the speaker in the belief that the listener would interpret the information using the same conventions. They are based on sensory capabilities or are learned from experience with the world and the language or by adopted standards, informal or formal. Though most of the mediate information will be stored in the minds of the speaker and listener, some of it may arrive contemporaneously with the utterance, such as situational information of a non-linguistic variety in the example.

Shannon's theory of communication – as Warren Weaver pointed out – "at first seems disappointing and bizarre [because it] "has nothing to do with meaning" and the measure it provides counter-intuitively links information with uncertainty. But for what we are calling potential information, Shannon showed the limitations of the physical channel and also how to use that channel to communicate within those limitations. He also dealt with disruption and corruption ("noise") of the potential information and how to cope with those problems (at a cost in efficiency, by adding redundancy). The mathematical theory of communication is recognized as a very important scientific contribution and communication engineers use techniques based on it routinely.

We need a similarly useful theory that deals with the delivery of "meaningful information", and since we already have Shannon's theory for potential information, we need to approach the mediate information. The new theory of mediate information must also deal with how to cope with noise, which may be more complex in the case of meanings that it is for Shannon information. It is clear that redundancy still plays a role, as humans typically have multiply connected concepts for any given word that they hear, and often multiple possible interpretations of the structure of a given string of potential information.

[2] Figure 2 is at the end of the paper.

3.3 Ontologies as Mediate Information

Figure 2 gives some idea about how the utterance, which is physically received and contains potential information, must be interpreted by mediate information shared by the source person and the destination person, but it really only scratches the surface. What is this "human knowledge" that is referred to and has its analogue in other organisms and in artificial information systems? If we wish to be very general about a system, we can work with its *ontology*, which should be, in our view, everything that the person has that will interpret inputs from language, from sensors, etc. There has been a lot written in recent years on the topic of an individual system's ontology and what it may contain, and we will not get into that directly in this paper.

In another paper, soon to be published [Reeker 02] it is argued that the needed extended ontology (or worldview) for complex systems is in general much more extensive than the ones that we see in the literature. It is argued that ontologies are inherently different within different individual organisms and yet the organisms (like the two humans in Figure 2) work together by making assumptions that are approximately correct in most instances. They also seek new knowledge if they have a feeling that they lack essential knowledge or feel that they do not understand or are not being understood. Traditional hierarchies or lattices of object classes, often called ontologies [Sowa 99], must be strengthened or extended for purposes of a scientific theory of knowledge and intelligence and the practical engineering consequences of such a theory. The notion of linking classification to sensory processes ("grounding") or to linguistic terms that are so grounded is essential, but not enough. Models that include explicit processes must be integrated with the ontology, not swept under the carpet as programs or parts of a knowledge base separate from the ontology. Each process in which an object is a participant can partially define the object. This means that the task of discovering (in organisms) or developing (in artifacts) an adequate worldview for utilitarian purposes must be more exacting than is sometimes implied. The additional burden will not go unrewarded, however, as it can improve the ability to engineer and evaluate intelligent systems, to automatically integrate systems, and to understand and control system behavior.

What this all says is that the "fabric of knowledge" is held together by a rich system of links, and communicating people can usually find some common links from their knowledge to whatever they hear from the people they are communicating with. The notion that the power of a

system for expressing mediate information is in all of the links between concepts and not just in the hierarchical nature of the system is not a new one (see Woods [75]) There is evidence in human cognitive processes that each action in which an entity performs may modify its meaning. Perhaps the strongest argument for this extended ontology need is the nature of science, where there has evolved a "fabric" of linked concepts that is shared by millions of people with a good deal of consistent understanding. The multiple connections and extensibility of the linked concepts of that fabric is widely considered to be a major strength of scientific theory.

3.4 Techniques for Practical Integration and Control

The use of information techniques to actually improve the integration of complex information systems for understanding and control is still a research topic. There are some approaches that appear hopeful, however, and these will now be discussed. We have not mentioned human intervention directly (which is the sole satisfactory method today), but human interaction may be involved in any of these techniques or combinations thereof.

3.4.1 The State Comparison Approach. The ideas described in section 3.1 above provides a clue to a method of looking at two subsystems and checking them with respect to their performance on given sets of data. If the information comprising the initial states of the two is the same and the information given as inputs is the same, then one can sometimes prove that the information given as outputs and the final states will also contain the same information. There may even be differences in the middle states of the algorithms or the number of intermediate states, but that does not matter in terms of the information ultimately produced. Making sure that this is true is clearly a strong requirement, and it may not always be provable either true or false. But the technique may be helpful in determining how the information is utilized and transformed (discussed in [Reeker, 1980]). It may be especially interesting in conjunction with some of the next three suggestions.

3.4.2 The "Work Analogy". The information measures that we have claimed to be inadequate for meaningful information do still have properties that we would suppose any information measure would have. The most important of these is the insight that information requires organization. If there were no organization in the world, then we have, for sure, what William James called a "booming, buzzing confusion". In fact, we would not even know that it was booming and buzzing because we would have no ordered way of retrieving meaning, let alone learning the words or their meaning in the first place. Which leads us to the idea that we do learn things and that learning is a form of organizing (of which, more in the next section). Thermodynamics tells us that organization takes energy, as the entropy principle is always spreading disorganization. Energy can be stored as "potential energy" that is just waiting for a force field to let it turn into kinetic energy.

The interesting thing here is that a force field has direction, so the kinetic energy released will be causing work to be done in that direction. Only along the direction of the force, which has a vector quantity, does that particular work get done. The rest of the energy is dissipated in some other ways, without necessarily doing any useful work. Is it possible, one might ask, to express the measure of meaning in an ontology through a set of vectors?

The reason that we are calling this the "work analogy" is that potential information can be made into knowledge (meaningful information) by its transformation through mediate information, which can be compared to a force field (where the dimensions are computed merely by the three geometric directions, as a cosine function of the force). Unfortunately, if we take that view, we come right back to the fact that we have too many dimensions in any vector that might possibly be broad enough to handle all of information. So does the work analogy work? It might, if supported by standard definitions of some dimensions.

There is some work ongoing already on putting together a common upper ontology, that could be extended to lower levels for specialized areas [Standard Upper Ontology (SUO) Working Group 02] Suppose that standard were to give us N orthogonal (or at least forming a vector space) information parameters. Then we might derive some analogy of the modern definition of work that treats it as taking place in the direction of each of these parameters, as a measure of meaningful content. It is hard to see how that would help us, since we are left with a space of arbitrary dimension. Under the circumstances, that makes the work analogy a problem, rather than a solution.

The psychologist and communication scholar Charles E. Osgood developed a measurement of a type of meaning (meaning being information content in much the same way that work is energy directed by a force) called connotative meaning. Connotative meaning is related to an individual's personal ontology [Osgood, 57] because it includes "shades of meaning" that may not be shared through a strict definition. The connotation is intended only

to be partial meaning, to be coupled with the more explicit denotation for a full definition in a particular context. In trying to measure it, Osgood postulated three dimension types or factors, within which pairs of adjectives would indicate denotations:

- Evaluative factor (example: good - bad)
- Potency factor (example: strong – weak)
- Activity factor (example: active - passive)

Osgood then measured each pair, for each factor, on a seven point Likert scale. He then constructed an n-dimensional space, n being the number of adjective pairs, for his “semantic differential”.

Clearly, much more than the semantic differential is needed to do the evaluation that can lead to integration of several manufacturing systems or bioinformatics systems. However, Osgood’s ideas fit into the idea of fuzzy frameworks, and it was an important step in trying to formalize the idea of how the vocabulary of humans may vary. Vocabulary, while not the same as ontology, is closely linked, and provides a way to get at human ontologies. So in a sense, it reflects the ontology in an approximate way. The possibility presented by this type of approach is most likely to be determination of closeness of various concepts by comparing dimensions based on a standard ontology with certain standardized dimensions as a major. It is not clear what value a unified measure based on some sort of standard dimensions for all ontologies would have, or how such standard dimensions would be defined.

3.4.3 Machine learning. Machine learning is becoming an important area in data and knowledge management, because it can potentially allow the development of enormous knowledge bases from enormous amounts of data that would not be economical or feasible for manual human development and because it is the basis of the field of data mining (along with data visualization, which allows humans to participate in the mining). As a short summary, there are three basic categories of machine learning generally recognized: unsupervised, supervised, and reinforcement learning. The one that requires the least detailed input information -- merely a similarity space in which the data are shown and the dimensions are attributes of the data – is unsupervised learning. If one had a list of words classified by Osgood’s semantic differential, then one could use unsupervised learning to cluster them in ways that reflected their denotational similarity. Clearly, the same could be done with concepts in an ontology based on attributes.

Supervised learning can actually learn to recognize things that exhibit a certain set of attributes, which are related in particular ways. It can do this even in cases where people have a hard time coming up with a computer program to recognize those things. An example is the astronomical phenomena that a particular astronomer may want to study. The program is given examples of things that exhibit a given phenomenon and examples of things that do not (preferably, some things that could be confused with things that exhibit the phenomenon but do not). It then looks at sky surveys, with their trillions (or more) of objects and finds a set of those which appear to exhibit the phenomenon. If taught well, such a program can be quite helpful to the astronomer, though it might make some mistakes (both false positives and false negatives), so it needs to be checked.

Reinforcement learning does not have to have all the examples, but it needs to have conditions that are rated “right” or “good” (which it will reward with positive numerical amounts) or are “wrong” or “bad” (which it will punish with negative numerical amounts). It is based on one model of animal conditioning. An example is a game-playing program that has been rewarded for good moves and punished for bad ones. The computer program TD-Gammon, which is probably the best backgammon program in the world, learned by reinforcement learning.

Statistical regression is another type of learning that can be programmed into a machine, and neural net models can also be used. Whatever type(s) of machine learning are chosen, the point is that a subsystem integrated into a complex system for something like supply chain management may be able to learn aspects of the behavior of other subsystems. These might include ontologies and state patterns, developing mediate information of value in informational interactions, “self-adapting” to the other subsystems in an integration process. Alternatively, a learning program could be used to find problems in the operation of full complex system – and maybe (through reinforcement) to alleviate the problems.

There is one more technique in human learning that has received a lot of attention in machine learning but has proven hard to implement in practice. That is analogy (or case-based reasoning). Humans use it regularly in what is called “transfer of learning”. As mentioned earlier in this paper, an informational objects in an ontology will have many activities and certain other informational objects linked to it. These will have certain attributes. Having that informational object and all of those links in his or her knowledge base, the person is

often capable of using an "approximately structurally identical" informational object in a new but somehow similar situation. This sort of transfer is something that might help in integrating similar systems and predicting their behavior.

In control, learning algorithms can be considered as optimization algorithms. But we may not need to optimize for systems to do what we want them to do. Herbert Simon, whom we have mentioned, and his observations on humans and human organizations provide the clue to another method that needs to be explored and the last one that we will mention here.

3.4.4 Satisficing. As a final suggestion for research directions, we turn again to the manner in which people settle perceived differences in their ontologies, using dialog to come to an approximate compromise, which is often a "near-enough" joint understanding. This is a sort of "satisficing", one of the important ideas stressed by Simon [57]. It is embodied in his statement,

> It appears probable that, however adaptive the behavior of organisms in learning and choice situations, this adaptiveness falls far short of the ideal "maximizing" postulated in economic theory. Evidently, organisms adapt well enough to "satisfice"; they do not, in general, "optimize".

Herb Simon was awarded the Nobel Memorial Prize for Economics in 1978 largely on this observation of "bounded rationality", backed up by empirical data and theories that have supplied models for many areas of science and are being explored carefully today as a solution for intractable computational problems [Zilberstein 97].

Although we are still trying to determine just how Simon's idea of "satisficing" would be used for practical integration, it fits well with the use of fallible learning, as described in §3.4.3, and as suggested by Simon's quotation above.

An encouraging thing about the work going on in satisficing presently is that it can provide some types of estimates of the bounds of rationality, where it has to be bounded to solve the problem in a reasonable time. We note that satisficing is being implemented as approximate reasoning, approximate modeling, optimal meta-reasoning, bounded optimality, and combinations of all these traits. In a situation where we are trying to compare ontologies which are governed by the activities in which they partake or attributes that need to be calculated and that may use different programs describing the activities or calculating the attributes, we are on the edge of undecidability (the general equivalence of two programs in undecidable). But it may be possible to decide if they are close enough to being the same to make them interoperable.

4. Summary

We do not yet know how to practically characterize complex systems in ways that allow the prediction of their behavior for purposes of optimal control. The traditional methods that have been used, even clever statistical methods that can handle limited indeterminacy, break down under the complexities that arises in supply chain management and other big problems for which the development of systems requires the integration of several complex subsystems and which evolve inevitably with time. These systems are informational in nature, or are hybrid physical-informational systems, in which the many informational dimensions add complexity not readily handled by the traditional approaches. We are looking at four approaches that might work together with one another, and also with the traditional approaches. One is replacing optimization with satisficing in some of the techniques; another is further exploring observations of states that do the same things. A third is pursuing programs that can define numbers of dimensions that allow mediate information to be described in vectors that can be controlled. The fourth uses machine learning by subsystems of the performance of other subsystems. Together, we hope these will give us better tools for handling complex informational systems like supply chain management.

5. References

Herrin, G. [01], Chaos Theory Applied To Manufacturing, *MMS Online*, Gardner Publications, Inc. (http://www.mmsonline.com/articles/0499cim.html).

Jaynes, E. T. [68] Prior Probabilities, *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, 221-241, 1968.

Jones, A., L. Reeker and A. Deshmukh [02], On information and performance of complex manufacturing systems, *Proceedings of 2002 Conference of Manufacturing Complexity Network*, Cambridge, UK, ed. G. Frizelle and H. Richards, 173-182, 2002.

Information: A Key to Supply Chain Performance A. Jones, L. Reeker and A. Deshmukh, World Automation Congress 2002, Orlando, FL, ISOMA-025.

Miller, G. A. and N. Chomsky [63], Finitary Models of Language Users, Chapter 13 of *Handbook of Mathematical Psychology*, vol. 2, Wiley, 1963.

Osgood, C. E., G. Suci and P. H. Tannenbaum [57], *The Measurement of Meaning*. Urbana: The University of Illinois Press, 1957. Reeker, L. [79], Mathematical foundations of computer science: a sampler, *Australian Computer Science Communications*, **1**, 4, 251-279, 1979.

Reeker, L. [00], Theoretical Constructs and Measurement of Performance and Intelligence in Intelligent Systems, in Meystel and Messina ed, *Measuring the Performance and Intelligence of Systems: Proceedings of the 2000 PerMIS Workshop, August 14-16, 2000*, NIST Special Publication 970.

Reeker, L. [02], The ontology: more than meets the sensors, forthcoming, (talk delivered at Theoretical Fundamentals of Intelligent Systems: Computational Semiotics, part of the Joint Conference on Intelligent Systems (JCIS); available from author).

Reeker, L. and Jones A. [01], Measuring the Impact of Information on Complex Systems, in Meystel and Messina ed, *Measuring the Performance and Intelligence of Systems: Proceedings of the 2001 PerMIS Workshop, September 4, 2001*. NIST Special Publication 982.

Shannon, C. and Weaver W. [71], *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1971.

Simon, H. A.. [57], *Models of Man: Social and Rational*, John Wiley, New York, NY, 1957.

Simon, H. A. [69], *The Sciences of the Artificial*. Third edition, Cambridge, MA, MIT Press, 1996. [First edition published 1969].

Sowa, J. F. [99], Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., 1999.

Standard Upper Ontology (SUO) Working Group [02], see http://suo.ieee.org.

Woods, William A. [75], "What's in a Link: Foundations for Semantic Networks". In D. Bobrow and A. Collins (eds.), Representation and Understanding: Studies in Cognitive Science, New York: Academic Press, 1975.

Zilberstein, S. [97], Formalizing the Notion of "Satisficing", AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning, Stanford, California, 1997.

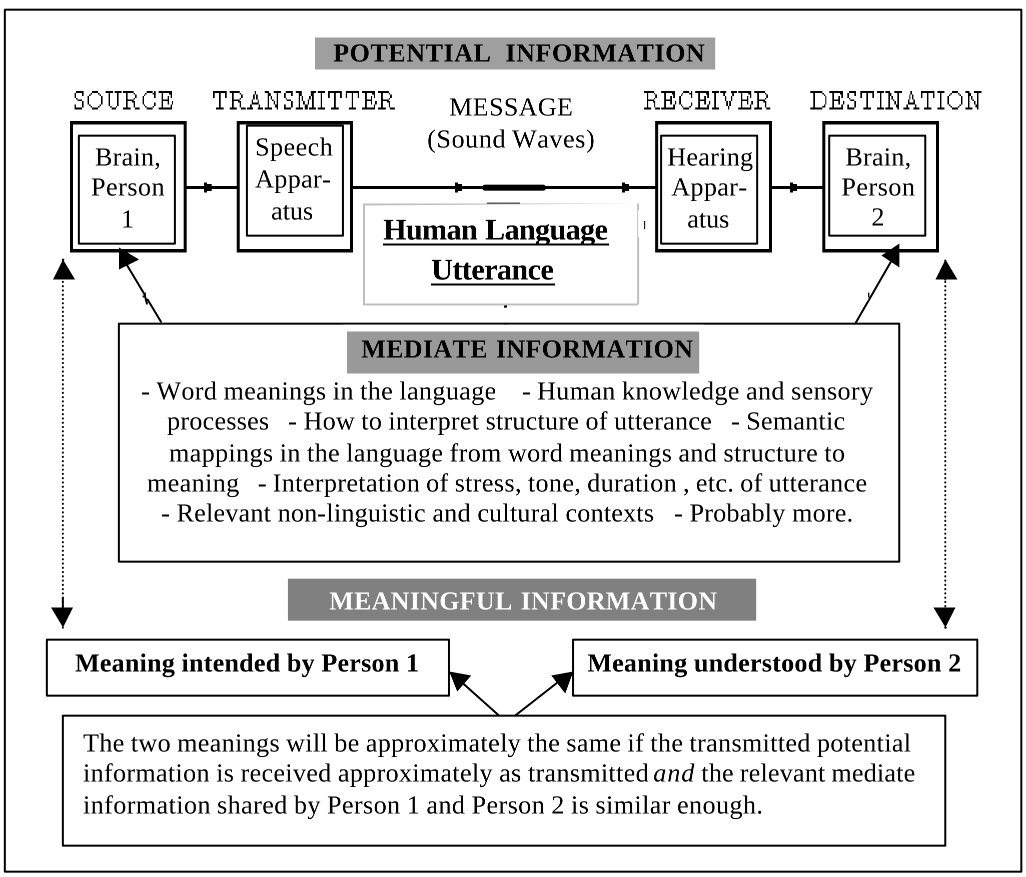

Fig. 2. Potential and Mediate Information Together are Needed to Convey Knowledge (Meaningful Information).